CHAPTER 2

LITERATURE REVIEW

2.0 History of Transistor

A transistor is a semiconductor device that uses a small amount of voltage or electrical current to control a larger change in voltage or current. The transistor is the fundamental building block of the circuitry that governs the operation of computers, cellular phones, and all other modern electronics. Because of its fast response and accuracy, the transistor may be used in a wide variety of digital and analog functions, including amplification, switching, voltage regulation, signal modulation, and oscillators. Transistors may be packaged individually or as part of an integrated circuit chip, which may hold thousands of transistors in a very small area.

*"We have called it the Transistor ,*T-R-A-N-S-I-S-T-O-R, *because it is a resistor or semiconductor device which can amplify electrical signals as they are transferred though if from input to output terminals. It is if you will, the electrical equivalent of vacuum tube amplifier. But there the similarity ceases. It has no vacuum, no filament, no glass tube. It is composed entirely of cold, solid substances." John Bardeen and Walter Brattain,1947* [1].

On 16 December 1947 William Shockley, John Bardeen and Walter Brattain succeeded in building the first practical point-contact transistor at Bell Labs. This first transistor, a point contact type, is shown in **Figure 2.1** [2]. In 1950 Shockley developed a radically different type of solid-state amplifier which became known as the Bipolar

Junction "transistor". Although it works on a completely different principle to the point-contact "transistor", this is the device which is most commonly referred to as a "transistor" today. This invention was the spark that ignited a huge research effort in solid state electronics. Bardeen and Brattain received the Nobel Prize in Physics, 1956, together with William Shockley, "for their researches on semiconductors and their discovery of the transistor effect." Shockley had developed a so-called junction transistor, which was built on thin slices of different types of semiconductor material pressed together. The junction transistor was easier to understand theoretically, and could be manufactured more reliably.
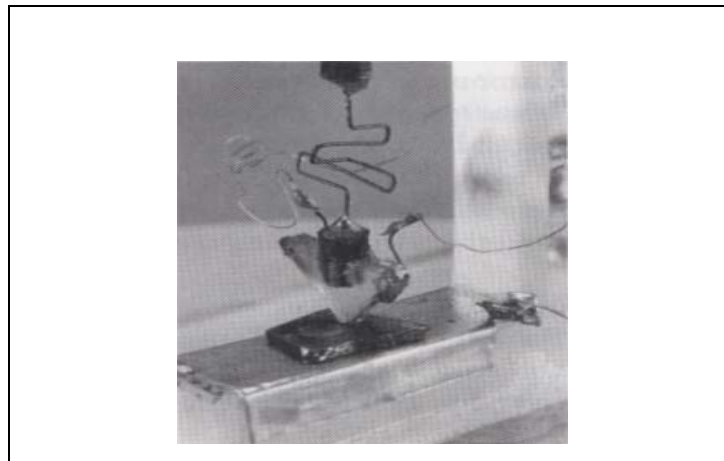

Figure 2.1. The first transistor made in Bell Laboratory. *Source*: Property of AT &T Archives [2].

The integrated circuit era arrived in 1958, at a workshop commemorating the tenth anniversary of the invention of the transistor in Bell Labs. Jack Kilby, a newly hired engineer at Texas Instruments, noticed that most discrete devices such as resistors, capacitors, diodes, and transistors, could be made from a piece of semiconductor material like silicon. Therefore, it would be possible to make them on the same piece of semiconductor substrate and connect them together to form a circuit. This could make a much smaller circuit, reducing the cost of the electronic circuit. As a new hire, Kilby didn't have any vacation time. So, while most of his colleagues were on summer vacation, Kilby consolidated his idea for an integrated circuit in the deserted R&D lab. When the other

workers returned from vacation, he presented his idea and started to demostrate it. Since no silicon substrate was readily available, Kilby used whatever he could find; namely a strip of germanium with one transistor already built on it. He added a capacitor and used the germanium bar itself to form three resistors. By connecting the transistor, capacitor, and three resistors from that half-inch long germanium bar with fine platinum wires, Jack Kilby made the first integrated circuit (IC) device, shown in **Figure 2.2**. At Texas Instruments, IC devices were called "bars" instead of "chips" or "dies" because of the shape of Jack Kilby's first IC device [2].
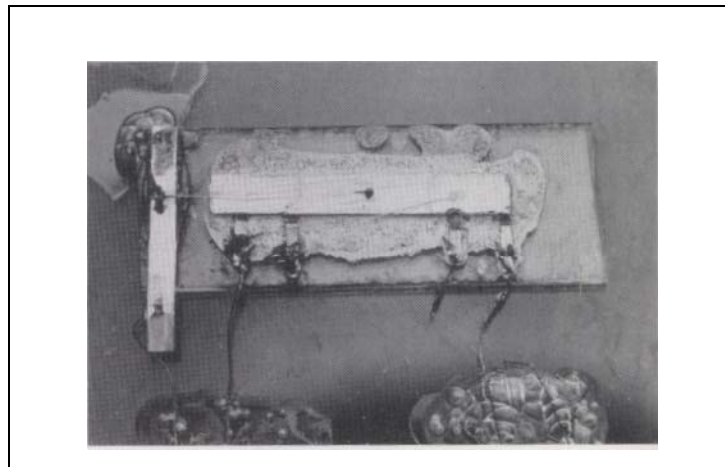

Figure 2.2. The first integrated circuit chip (bar) made by Jack Kilby. *Source:* Photo courtesy of Texas Instruments [2].

The invention of the transistor earned the Nobel Prize in Physics in 1956 for Bardeen, Brattain, and their co-worker William Shockley. Kilby received the Nobel Prize in Physics in 2000 for the invention of the integrated circuit [1].

2.0.1. Metal – Oxide – Semiconductor – Field – Effect Transistor (MOSFET)

"The silicon transistor metal-oxide-semiconductor field-effect transistor (MOSFET or MOS transistor) did not become significant commercially until two decades after the 1948 announcement of the invention of the transistor by Bell Laboratories. The underlying concept of the MOSFET--modulation of conductivity in a semiconductor triode structure by a transverse electric field--first appeared in a 1928 patent application. It was confirmed experimentally in 1948. However, early devices were not practical due to surface problems. Although these were solved at Bell Laboratories in 1958, Bell remained committed to earlier transistor technology. Development of the 'other transistor' was first pursued elsewhere. It was finally the needs of computers and the opportunities created by integrated circuits that made the silicon MOSFET the basic element of late 20th-century digital electronics" [3].

The basic concept of an insulated-gate field-effect transistor (IGFET) is shown in **Figure 2.3**. The IGFET has been the most common type of field effect transistor from the first commercial silicon MOSFETs to the MOS technology later used in integrated circuits. It depends upon the modulation of the current in a conducting semiconductor channel by a transverse electric field due to a gate electrode that is insulated from the channel. The insulating layer in an IGFET is analogous to the vacuum in a vacuum tube. In a properly biased vacuum tube there is no current in the circuit of the grid, which is thus akin to the IGFET's gate.

If the source and the p-type silicon substrate are grounded, if the drain is at a positive potential, and if the gate is positive and exceeds the device's threshold potential, an inversion layer of mobile electron will form a conducting channel in the gated region adjacent to the upper surface of the substrate. An insulating layer of silicon dioxide separates the gate from the conducting channel. Electron enter through the source, leave through the drain and are subject to the controlling action of the gate in the region between the heavily- doped n-type junction islands. The larger the gate voltage, the larger will be the number of electrons at the silicon surface and the greater the conductance of the inversion

layer. The silicon substrate can be either p-type, leading to the n-channel (electron conduction) devices shown, or n-type, leading to a p-channel device, in which holes are carries, p-type junction island are, of course, required in n-bulk devices and the drain and gate voltage are negative, rather than positive, with the source and substrate grounded. Note that the voltage applied to drain in these MOSFETs is of a polarity to reverse bias the diffused p-n junction at the drain contact.
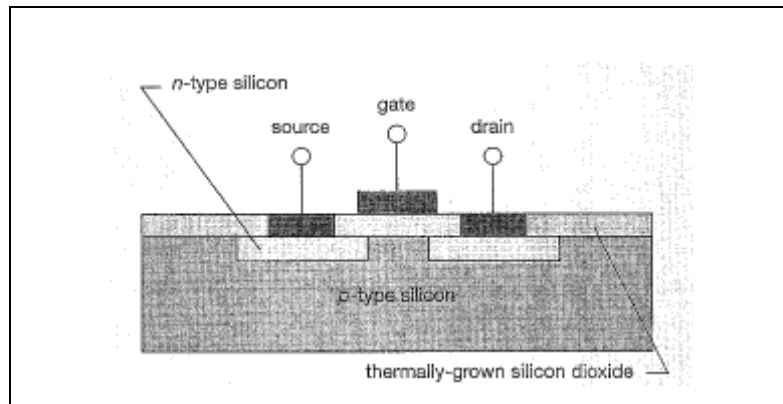

Figure 2.3. Schematic cross-section of an n-channel enhancement-mode silicon metal-oxide semiconductor field-effect transistor [2].

2.1   Operation MOSFET Transistor

The MOS transistor is a majority-carrier device in which the current in a conducting channel between the source and drain is controlled by voltage applied to the gate. In an ammos transistor, the majority carriers are electrons and in a pMOS transistor, the majority carriers are holes.

2.1.1 NMOS Transistor

An NMOS structure is shown in **Figure 2.4**. It has a conducting gate, silicon substrate, and a thin layer of silicon dioxide sandwiched in between. For an NMOS, the substrate is p-type, and the source and drain are heavily doped with n-type dopant (which is why they are noted as n+). Source and drain are symmetric; normally, the grounded side is called the *source* and the biased side the *drain.*
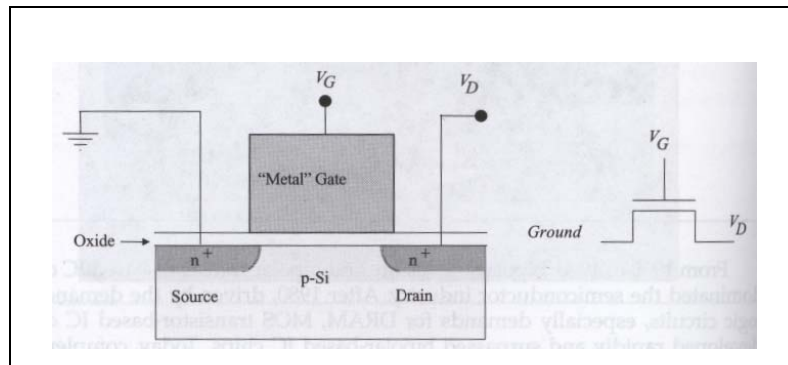

Figure 2.4. An NMOS transistor structure.

When no bias voltage is applied to the gate, no matter how the source- drain is biased, no current can go through from source to drain, or vice versa when the gate is biased positively, it will generate positive charges at the metal-oxide surface. The gate's silicon dioxide is a thin dielectric layer between metal and semiconductor. Like a capacitor, the positive charge on the metal-oxide surface will expel positive charges (holes, majority carriers) from the silicon-oxide surface and attract negative charges (electrons, minority, carriers) to that surface [2]. When the gate voltage is higher than the threshold,' voltage, V_G > V_T > 0, the silicon-oxide surface will accumulate enough electrons to form a channel and allow electrons from source and drain to flow across it. That is why an NMOS is also called an n-channel MOSFET, or NMOSFET. This process is illustrated in **Figure 2.5**. By controlling the gate voltage, the electric field will affect the conductivity of the semiconductor device and switch the MOS transistor on and off. This is the reason it is call a field-effect transistor (FET).
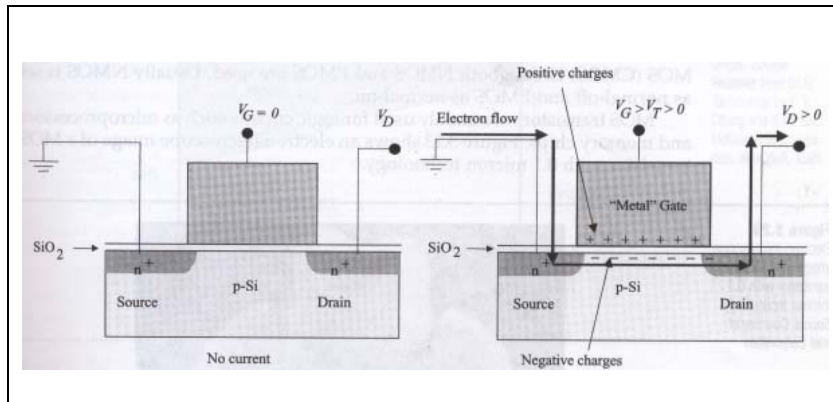
Figure 2.5. Switch process of the NMOS.

2.1.2 PMOS Transistor

For a PMOS, the substrate is an n-type semiconductor, and the source/ drain are heavily p-type doped. It uses negative gate bias to generate negative charges (electrons) on a metal-oxide surface, which in turn drives away electrons (majority carriers of substrate) and attracts holes (minority carriers) to the silicon-oxide surface to form a hole-channel underneath the gate. Holes from source and drain then can flow through the channel and conduct electric current between source and drain (**Figure 2.6**) [2]. With positive bias, the PMOS transistor turns off.
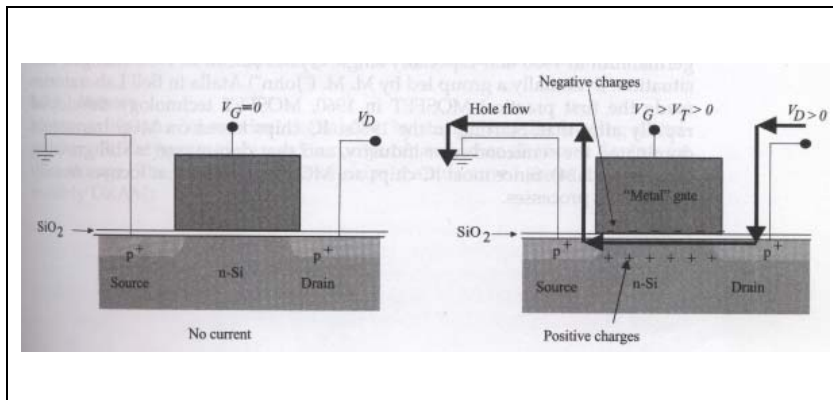

Figure 2.6. Structure and switch process of the PMOS.

2.2   Methodology Process Fabrication of Transistor

In the early stages of the IC industry, most Tabs made bipolar transistor-based IC chips. When they made MOSFET-based IC chips, they used p-channel MOSFET or PMOS because of technology limitations. For exactly the same design (same gate material, geometry, and substrate and source/drain dopant concentration), NMOS is significantly faster than PMOS, because electrons have two to three times higher mobility than holes. However, difficulties associated with the NMOS process could not be solved at that time without ion implantation technology. It was much easier to make PMOS than NMOS using diffusion techniques for silicon doping. The PMOS process flow using 1960s technologies is shown in **Table 2.1**. It used blanket-field oxide for isolation, boron diffusion for source/drain doping, and aluminum-silicon alloy for gate and interconnection. About 1 % silicon was added into the aluminum to saturate the silicon dissolving, which prevented the junction spiking of the aluminum caused by the silicon dissolution. The minimum feature size was about 20 µm.

Table 2.1 PMOS Process sequence [2].

Wafer clean	(R)	Etch oxide	(R)
Field oxidation	(A)	Strip photoresist	(R)
Mask 1. (Source/drain)	(P)	Aluminum deposition	(A)
Etch oxide	(R)	**Mask 4**. (Metal)	(P)
Strip photoresist/clean	(R)	Etch aluminum	(R)
S/D diffusion	(A)	Strip photoresist	(R)
(boron/oxidation)	(P)	Metal anneal	(H)
Mask 2. (Gate)	(R)	CVD oxide	(A)
Etch oxide	(R)	**Mask 5**. (Bonding pad)	(P)
Strip photoresist/clean	(A)	Etch oxide	(R)
Gate oxidation	(P)	Test and Packaging	
Mask 3. (Contact)			
Key: A = adding, H = heating, P = pattering, R = removing process.			

The whole PMOS process has five mask steps. Each is a photolithography step, including: wafer clean, prebake, primer application and photoresist coating, soft bake, alignment and exposure, development, pattern inspection, and hard bake.

2.2.1   Mask and Reticles

Semiconductor manufacturing entails the formation of various patterns on wafers. These patterns define the structure of and interconnection between the different components and features of the integrated circuit. The patterns are formed on wafers using patterning tools known as masks and reticles. A mask is defined as a tool that contains patterns which can be transferred to an entire wafer or another mask in just a single exposure. A reticle is defined as a tool that contains a pattern image that needs to be stepped and repeated in order to expose the entire wafer or mask.. The 'polarity' of a mask or reticle can either be positive or negative. A positive mask or reticle has background areas (or fields) that are clear or transparent, which is why a positive mask or reticle is also known a 'clear-field' tool.

A negative mask or reticle has fields that are opaque, which is why a negative mask or reticle is also known a 'dark-field' tool. There are many ways by which a pattern may be transferred to a wafer using a mask, a reticle, or a combination of both. Regardless of the pattern transfer process, everything starts with a set of pattern data that are converted into an actual pattern by a 'pattern generator.' Commonly-used pattern generators include: 1) plotters; 2) optical pattern generators; and 3) electron beam pattern generators. The patterns formed on a reticle can be transferred directly onto the wafer, or they may first go to a mask which is the one that transfers the patterns to the wafer. Patterns on masks generally get transferred to the wafer directly.

2.2.2 Thermal Oxidation

Oxidation is one of the most important thermal processes. It is an adding process, which adds oxygen to a silicon wafer to form silicon dioxide on the wafer surface. Silicon is very reactive to oxygen; thus, in nature most silicon exists in the form of silicon dioxide, such as quartz sand: It reacts with oxygen very quickly, and forms silicon dioxide on the silicon surface. The reaction can be expressed as:

$$Si + O_2 \longrightarrow SiO_2 \tag{2.1}$$

Silicon dioxide is a dense material that fully covers the silicon surface. To continue the oxidization of silicon, oxygen molecules have to diffuse across the oxide layer to reach the silicon atoms underneath and react with them. The growing silicon dioxide layer increasingly blocks and slows the oxygen. When bare silicon is exposed to the atmosphere, it reacts almost immediately with oxygen or moisture in the air and forms a thin layer (about 10 to 20 Å) of silicon dioxide, called *native oxide*. The thickness of the native oxide is enough to stop the further oxidation of the silicon at room temperature. Figure 2.7 illustrates the oxidation process.

In the oxidation process, oxygen is in the gas phase, and the silicon is from the solid substrate. Therefore, while silicon dioxide is growing, it consumes the substrate silicon, and the film grows into the silicon substrate. Oxygen is widely used in oxide formation processes such as thermal oxidation, CVD, and reactive sputtering deposition. It is also commonly used in etch and photoresist stripping processes.
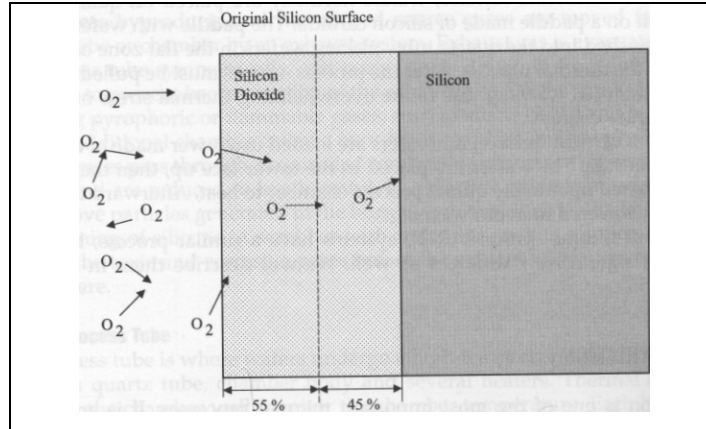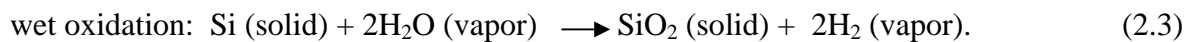
Figure 2.7. Silicon oxidation process [2].

At high temperatures, thermal energy makes oxygen molecules move much faster, which can drive them to diffuse across an existing oxide layer and react with silicon to form more silicon dioxide. The higher the temperature, the faster the oxygen molecules, and the quicker the oxide film grows. The oxide film quality is also better than that grown at lower temperature. Therefore, to get high-quality oxide film and fast growth rate, oxidation processes are always performed in a high-temperature environment, normally in a quartz furnace. Oxidation is a slow process; even in furnaces over 1000 °C, the thick oxide (> 5000 Å) still takes several hours to grow. The reactions for dry and wet oxidation are governed by the following equations:

dry oxidation: Si (solid) + O_2 (vapor) $\longrightarrow$ SiO$_2$ (solid). (2.2)

wet oxidation: Si (solid) + 2H$_2$O (vapor) $\longrightarrow$ SiO$_2$ (solid) + 2H$_2$ (vapor). (2.3)

Wet oxidation, when the oxidizing atmosphere contains water vapor. The temperature is usually between 900 °C and 1000 °C. This is also called *pyrogenic* oxidation when a 2:1 mixture of hydrogen and oxygen is used. Wet oxidation is a rapid process. Dry oxidation when the oxidizing atmosphere is pure oxygen. Temperature is in the region of 1200 °C to achieve an acceptable growth rate. Dry oxidation form a better quality oxide than wet oxidation. It is used to form thin, highly controlled gate oxides, while wet oxidation may be used to form thick field oxides.

2.2.2.1 Oxidation Rate

When oxygen starts to react with silicon, it forms a silicon dioxide layer, which separates the silicon atoms from oxygen molecules. When the oxide has just started to grow and the oxide layer is very thin (< 500 Å), oxygen molecules can penetrate the oxide with few collisions in the oxide layer and reach the silicon, to react and continue to grow the silicon dioxide film. This is called the *linear growth regime,* in which the oxide thickness is linearly increasing with the growth time. When the oxide film becomes thicker, oxygen molecules can no longer pass through the oxide layer without many collisions with the atoms inside the oxide film. They have to diffuse across the growing oxide to reach silicon and react with it to form silicon dioxide. This is called the *diffusion-limited regime;* the oxide growth rate is slower than in the linear growth regime. Figure 2.8 illustrates the two regimes. A and B in the equations in Figure 2.8 are the two coefficients related to oxide growth rate, which are determined by many factors, such as oxidation temperature, oxygen source (O_2 or H_2O), silicon-crystal orientation, dopant type and concentration, pressure.
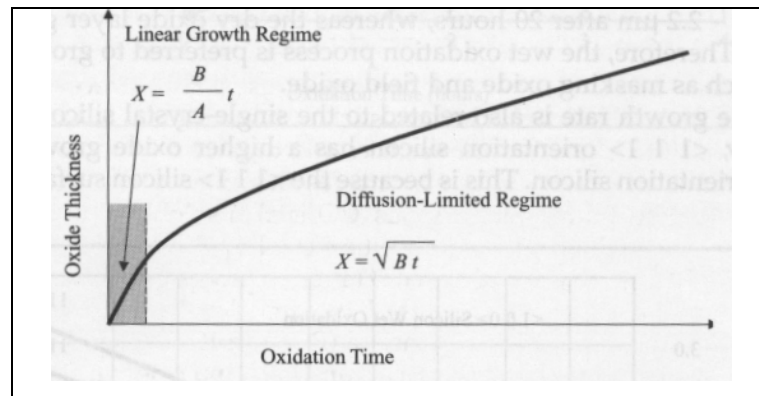

Figure 2.8.Illustrates of the two oxidation regimes [2].

Oxide growth rate is *very* sensitive to temperature, because the oxygen diffusion rate in silicon dioxide is exponentially related with temperature, $D \propto \exp(-E_a/kT)$. Here D is the diffusion coefficient, E_a is the activation energy, $k = 2.38 \times 10^{-23}$ J/K is the Boltzmann constant, and T is the temperature. Increasing temperature can significantly increase both B and $B/A,$ and the oxide growth rate.

Oxide growth rate is also related to the oxygen source. Dry oxidation with O_2 has a lower oxide growth rate than wet oxidation with H_2O. This is because the diffusion rate of the oxygen molecule O_2 in silicon dioxide is lower than that of the hydroxide HO generated from the dissociation of H_2O molecules at high temperature. The oxide growth rates of the dry and wet oxidation processes are illustrated in **Figures 2.9** and **2.10**.
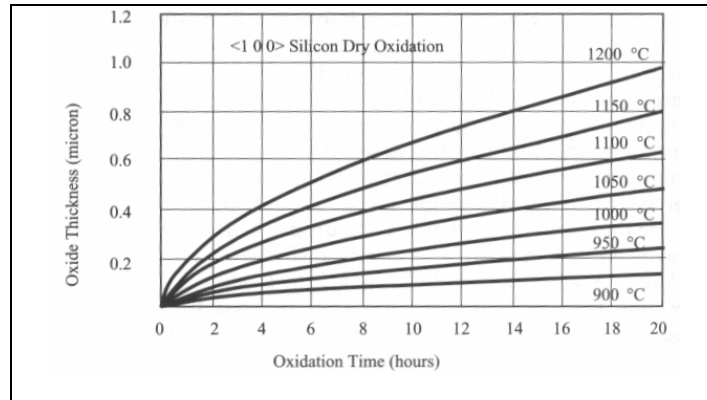

Figure 2.9 Dry oxide growth on <1 0 0> surface [2].

From the **figures 2.10** wet oxidation is significantly faster than dry oxidation. For example, with <1 0 0> silicon at 1000 °C, the wet oxide layer grows to - 2.2 µm after 20 hours, whereas the dry oxide layer grows to only 0.34 µm. Therefore, the wet oxidation process is preferred to grow thick oxide layers such as masking oxide and field oxide. Oxide growth rate is also related to the single-crystal silicon orientation. Normally, <1 1 1> orientation silicon has a higher oxide growth rate than <1 0 0> orientation silicon. This is because the <1 1 1> silicon surface has higher silicon atom density than that of the <1 0 0> silicon surface, and can provide more silicon atoms to react with oxygen and form a thicker silicon dioxide layer.

Compare **Figure 2.11**, which shows the wet oxidation rate of the <1 1 1> silicon, with Figure 2.10, which shows the wet oxidation rate of the <1 0 0> silicon. Oxidation rate is also related to dopant and dopant concentration. Generally, heavily doped silicon oxidizes faster than lightly doped silicon. During oxidation, boron in the silicon tends to be drawn up to the silicon dioxide and cause depletion of the boron concentration at the

silicon-silicon dioxide interface. N-type dopants such as phosphorus, arsenic, and antimony have the opposite effect. Oxide growth rate is affected by time, temperature, and pressure. More specifically, oxide growth is accelerated by an increase in oxidation time, oxidation temperature, or oxidation pressure. Other factors that affect thermal oxidation growth rate for SiO_2 include: the crystallographic orientation of the wafer; the wafer's doping level; the presence of halogen impurities in the gas phase; the presence of plasma during growth; and the presence of a photon flux during growth.
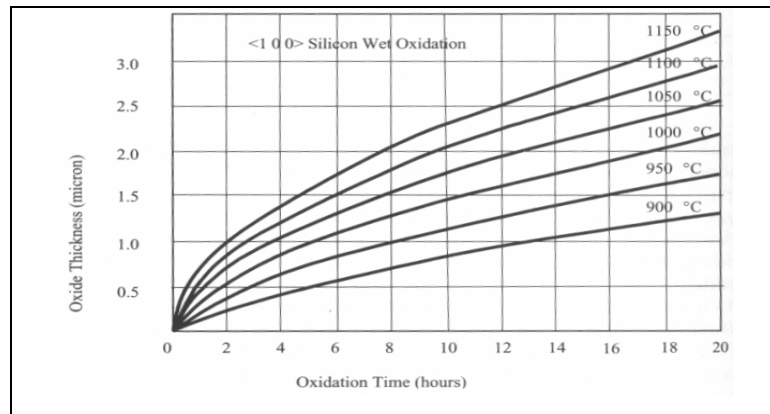

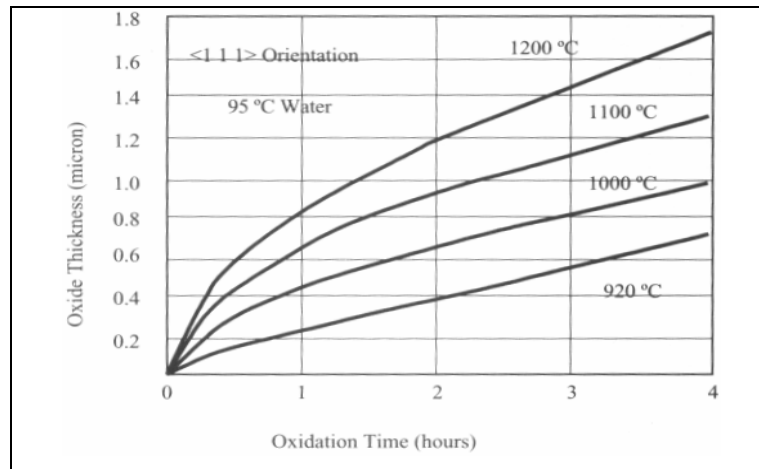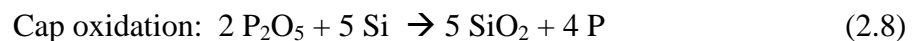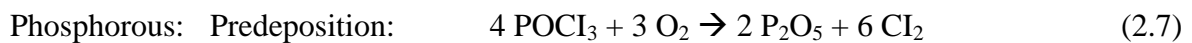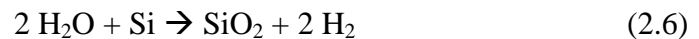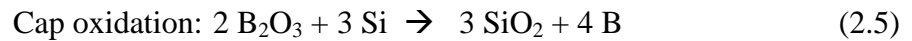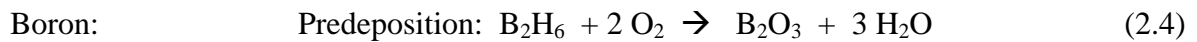**Figure 2.10**. Wet oxide growth on <1 0 0> surface [2].

Figure 2.11. Wet oxide growth on <1 11> surface [2].

2.2.3 Diffusion

Diffusion, the movement of a chemical species from an area of high concentration to an area of lower concentration, is one of the two major processes by which chemical species or dopants are introduced into a semiconductor (the other one being ion implantation) . The controlled diffusion of dopants into silicon to alter the type and level of conductivity of semiconductor materials is the foundation of forming a p-n junction and formation of devices during wafer fabrication. There are two major ways by which to deposit impurities into a substance by thermal diffusion. In the first method, known as predeposition, a flux of impurities continuously arrives at the surface of the substrate such that the concentration gradient of the impurity remains constant at the surface of the substrate. In the second method, known as redistribution or drive-in diffusion, a thin layer of the impurity material is deposited on the substrate. In this case, the impurity gradient at the surface of the substrate decreases with time.

Boron: $\qquad$ Predeposition: $B_2H_6 + 2\,O_2 \rightarrow B_2O_3 + 3\,H_2O$ $\qquad$ (2.4)

$\qquad$ Cap oxidation: $2\,B_2O_3 + 3\,Si \rightarrow 3\,SiO_2 + 4\,B$ $\qquad$ (2.5)

$\qquad$ $2\,H_2O + Si \rightarrow SiO_2 + 2\,H_2$ $\qquad$ (2.6)

Phosphorous: Predeposition: $\qquad$ $4\,POCI_3 + 3\,O_2 \rightarrow 2\,P_2O_5 + 6\,CI_2$ $\qquad$ (2.7)

$\qquad$ Cap oxidation: $2\,P_2O_5 + 5\,Si \rightarrow 5\,SiO_2 + 4\,P$ $\qquad$ (2.8)

2.2.4 Photolithography

Photolithography is the patterning process that transfers the designed pattern from the mask or reticle to the photoresist on the wafer surface. Photolithography was first used in the printing industry and has long been used to make printed circuit boards. It was adapted in the semiconductor industry for transistor and integrated circuit manufacture in the 1950s. It is the most crucial process step in IC fabrication, since the device and circuit designs are transferred to the wafer by either etch or ion implantation through the pattern defined on the photoresist on the wafer surface by the photolithography process.This

consists of using photoresist materials and masks to selectively expose or cover areas on the die to which new materials may be added or from which existing materials may be removed. Photolithography consists of a series of steps and, consequently, requires several individual equipment to accomplish these steps.

The equipment used in photolithography include: 1) resist coating equipment to deposit the photoresist on the wafer; 2) ovens for soft-baking the photoresist; 3) exposure systems to subject the resist to some form of radiation; 4) development systems to remove or retain (depending on photoresist type) the exposed areas of the resist, leaving behind a mask pattern that may be used for other wafer fabrication processes.

The photolithography process can be subdivided into three main operations: photoresist coating, alignment and exposure, and photoresist developing. First the wafer is coated with a thin layer of photosensitive material, photoresist (PR), which is exposed by ultraviolet light through a mask or reticle with the pattern of clear and dark areas generated by the plotter, based on IC design. The chemistry of the exposed photoresist will be changed by the, photochemical reactions under the clear areas where the UV light can pass through. For the positive photoresist, which is commonly used in advanced semiconductor Tabs, the exposed area is dissolved in the developer solution, leaving the unexposed PR on the wafer surface, which reproduces the dark pattern on the mask/reticle.

2.2.4.1 Photoresist

Photoresists are the photosensitive materials used to temporarily coat the wafer and to transfer the optical image of the chip design on the mask or reticle to the wafer surface. It is similar to the photosensitive material on the plastic of photo films, which can transfer the optical images focused by the lens of a camera on the plastic film surface. But unlike these photosensitive coatings on film, the photoresists are not very sensitive to the visible light, nor are they required to be sensitive to the changes of color, or *gray levels,* of the light.

Since it is mainly sensitive to ultraviolet (UV) light and insensitive to visible light, the photolithography process does not need a darkroom, as is required for film development. Since photoresists are insensitive to yellow light, all semiconductor fabs use yellow lights to illuminate photolithography areas. Photoresist layers have two basic functions: 1) precise pattern formation; and 2) protection of the substrate from chemical attack during the etch process. Typical resists consist of three components: 1) the resin, which serves as the binder of the film; 2) the inhibitor or sensitizer, which is the photoactive ingredient; and 3) the solvent, which keeps the resist in liquid state until it is processed. There are two kinds of photoresists, positive and negative. For the negative PR, the exposed parts become cross-linked and polymerized due to the photochemical reaction, which hardens and remains on the wafer surface after development, whereas the unexposed parts are dissolved by the developer solution.

For the positive photoresist, the main component is novolac resin, which is a cross-linked polymer before the exposure. After the exposure process, the exposed part's cross-links break down and become "softened" due to the photochemical reaction called *photosolubilization,* and will be dissolved by the developer while the unexposed parts remain on the wafer surface.

Figure 2.12 illustrates the two different kinds of photoresists and their pattern-transfer processes. The image of positive photoresist is the same as the image on the mask or reticle, and the image of negative photoresist is the reversed image. The roll of photo film one purchases is usually a negative film; the images on a negative film after developing are the reversed images the camera has taken, which need exposure and developing again on the negative photo papers to print the normal images.
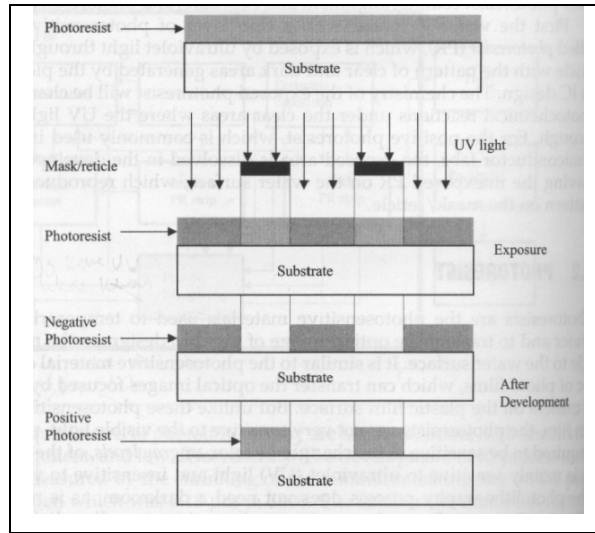
Figure 2.12. Patterning process with negative and positive photoresists.

For negative photoresist, the developer solution is mainly xylene. It dissolves the unexposed photoresist while some developer solvents can be absorbed in the exposed, cross-linked photoresist. It causes the PR "swelling" effect, which distorts the pattern features and limits the resolution to about 2 to 3 times the thickness of the photoresist. Negative photoresist was widely used in the semiconductor industry before the 1980s, when the minimum feature size was larger than 3 µm. It is no longer used in advanced semiconductor fabs because of its poor resolution. Positive photoresists do not absorb the developer solvents. They can achieve much higher resolution, and are widely used for the photolithography processes. **Figure 2.13** illustrates the resolution of the negative and positive photoresist.
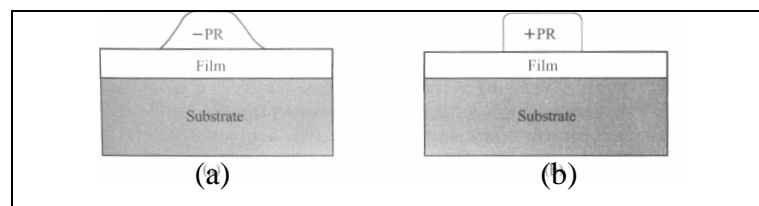

Figure 2.13. Photoresist (a) Negative and (b) Positive.

2.2.4.2 Photolithography Process

The photolithography process includes three major steps: PR coating, exposure, and developing. To achieve high resolution, photolithography also has lots of baking and chilling steps. For the older, all-manual process technology, the photolithography process flow had eight steps: wafer clean, prebake, spin coating primer and PR, soft bake, alignment and exposure, developing, pattern inspection, and hard bake. If a wafer failed to pass the inspection, it bypassed the hard bake step, the PR stripped, and the whole process repeated until it passed the inspection.

For the advanced photolithography process, the three basic steps are the same. Detailed steps are added to achieve high photolithography resolution. Integrated track-aligner systems are widely used to improve the process yield and throughput. Because all the coating, baking/cooling, exposure, and developing process steps are performed in a track-aligner system, the pattern inspection is performed after the hard bake. **Figure 2.14** shows a flowchart of the photolithography process.
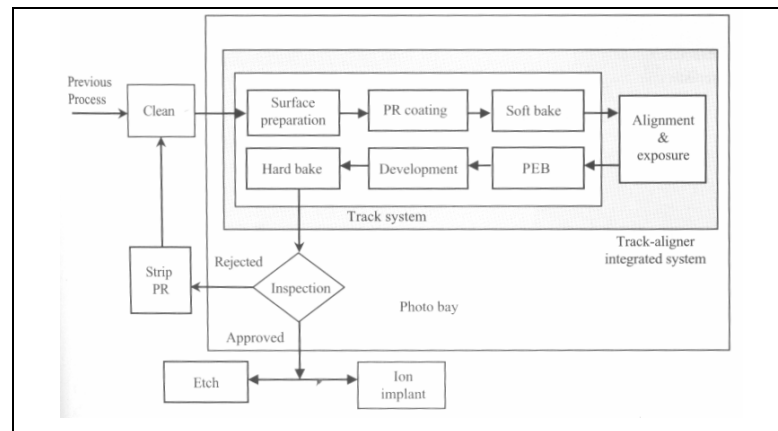

Figure 2.14. Flowchart of the Photolithography process.

Photoresist processing, or simply resist processing, basically consists of six steps: 1) dehydration and priming; 2) resist coating; 3) soft baking; 4) exposure; 5) development; and 6) post-development inspection. Prior to the application of resist to a wafer, the wafer must

be free of moisture and contaminants, both of which cause a multitude of resist processing problems. Dehydration baking is performed to eliminate any moisture adsorbed by substrate surfaces, since hydrated substrates result in adhesion failures. The bake is usually performed between 400 deg C to 800 deg C. Convection ovens may be used for baking up to 400 deg C, while furnace tubes are used for 800 deg C baking. After dehydration baking, the wafer is coated with a pre-resist priming layer designed to enhance the adhesion properties of the wafer even further. One of the most common primers used for this purpose is *hexamethyldisilazane* (HMDS) [1,2].

Resist coating, or the process itself of producing a uniform, adherent, and defect-free resist film of correct thickness over the wafer, is usually performed by spin-coating. Spin-coating consists of dispensing the resist solution over the wafer surface and rapidly spinning the wafer until it becomes dry. Most spin-coating processes are conducted at final spin speeds of 3000-7000 rpm for duration of 20-30 seconds.

Resist coating is followed by a soft bake, which is done to: 1) drive away the solvent from the spun-on resist; 2) improve the adhesion of the resist to the wafer; and 3) anneal the shear stresses introduced during the spin-coating. Soft baking may be performed using one of several types of ovens (e.g., convection, IR, hot plate). Soft-bake ovens must provide well-controlled and uniformly distributed temperatures and a bake environment that possesses a high degree of cleanliness. The recommended temperature range for soft baking is between 90-100 deg C, while the exposure time needs to be established based on the heating method used and the resulting properties of the soft-baked resist.

After a wafer has been coated with photoresist and subjected to soft baking, it has to undergo exposure to some form of radiation that will produce the pattern image on the resist. The pattern is formed on the wafer using a mask, which defines which areas of the resist surface will be exposed to radiation and those that will be covered. Alignment and exposure are the most critical steps of the photolithography process, this process determine the success of transferring for the IC design pattern on the mask or reticle to the photoresist on the wafer surface. The chemical properties of the resist regions struck by radiation

change in a manner that depends on the type of resist used. Irradiated regions of positive photoresists will become more soluble in the developer, so positive resists form a positive image of the mask on the wafer. Negative resists form a negative image of the mask on the wafer because the exposed regions become less soluble in the developer. Development, which is the process step that follows resist exposure, is done to leave behind the correct resist pattern on the wafer which will serve as the physical mask that covers areas on the wafer that need to be protected from chemical attack during subsequent etching, implantation, lift-off, and the like. The development process involves chemical reactions wherein unprotected parts of the resist get dissolved in the developer. A good development process has a short duration (less than a minute), results in minimum pattern distortion or swelling, keeps the original film thickness of protected areas intact, and recreates the intended pattern faithfully.

Development is carried out either by immersion developing, spray developing, or puddle developing. Regardless of method used, it should always be followed by thorough rinsing and drying to ensure that the development action will not continue after the developer has been removed from the wafer surface.

2.2.5 Etch

Etching is the process of removing regions of the underlying material that are no longer protected by photoresist after development. The rate at which the etching process occurs is known as the etch rate. The etching process is said to be isotropic if it proceeds in all directions at the same rate. If it proceeds in only one direction, then it is completely anisotropic. Since etching processes generally fall between being completely isotropic and completely anisotropic, an etching process needs to be described in terms of its level of isotropy. Wet etching or etching with the use of chemicals, is generally isotropic. On the other hand, dry etching processes that employ reactive plasmas are generally anisotropic. Silicon (single-crystal or poly-crystalline) may be wet-etched using a mixture of nitric acid

(HNO$_3$) and hydrofluoric acid (HF). The nitric acid consumes the silicon surface to form a layer of silicon dioxide, which in turn is dissolved away by the HF. The over-all reaction is as follows:

$$Si + HNO_3 + 6\ HF \rightarrow H_2SiF_6 + HNO_2 + H_2 + H_2O \tag{2.9}$$
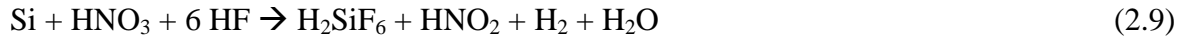
Silicon dioxide may, as mentioned above, be wet-etched using a variety of HF solutions. The over-all reaction for this is:

$$SiO_2 + 6\ HF \rightarrow H_2 + SiF_6 + 2\ H_2O \tag{2.10}$$

Water-diluted HF with some buffering agents such as ammonium fluoride (NH$_4$F) is a commonly used SiO$_2$ enchant formulation .Wet etching of aluminum and aluminum alloy layers may be achieved using slightly heated (35-45°C) solutions of phosphoric acid, acetic acid, nitric acid, and water. Again, the nitric acid consumes some of the aluminum material to form an aluminum oxide layer. This oxide layer is then dissolved by the phosphoric acid and water, as more Al$_2$O$_3$ is formed simultaneously to keep the cycle going.

2.2.6 Metallization

Metallization refers to the metal layers that electrically interconnect the various device structures fabricated on the silicon substrate. Thin-film aluminum is the most widely used material for metallization, and is said to be the third major ingredient for IC fabrication, with the other two being silicon and SiO$_2$.

Aluminum is very suitable for this purpose with its very low resistivity and its adhesion compatibility with SiO$_2$. A disadvantage of Al as the metallization material is its low melting temperature (660 °C) and the low Al-Si eutectic temperature (577 °C). These restrict the maximum processing temperature once the Aluminum layer has been deposited.

Actually, aluminum alloys (lightly doped Al) such as Al-Cu are preferred to pure aluminum for metallization because these inhibit problems like *electromigration* and junction spiking.

Al metal layers are usually can deposited either by *evaporated* or *sputtering*. Evaporated is performed by passing a high electrical current through a thick aluminum wire in vacuum chamber. Some of the aluminum atoms are vaporized and deposited on the wafer. An improved form of evaporated that suffers less from contamination focuses an electron beam at container of aluminum to evaporated the metal. Sputtering is achieved by generating a gas plasma by ionizing an inert gas using an RF or DC electric field. The ions are focused on an aluminum target and the plasma dislodges metal atoms, which are then deposited on the wafer.

Wet or dry etching can be used to remove unwanted metal. *Piranha solution* is a 3:1 to 5:1 mix of sulphuric acid and hydrogen peroxide that is used to clean wafers of organic and metal contamination or photoresist after metal pattering. Plasma etching is a dry etch process with fluorine or chlorine gas used for metallization steps. The plasma charges the etch gas ions, which are attracted to the appropriately charged silicon surface. Very sharp etch profile can be achieved using plasma etching.

2.2.7   Passivation

The final processing steps are to add a protective layer called *passivation* that prevents the ingress of contaminations. Opening in the passivation layer allow connection to I/O pads and test probe points. After passivation, further steps can be performed such as *bumping*, which allows the chip to be directly connected to a circuit board using plated solder bumps in the pad openings.

2.3. Lift – off Process

In semiconductor wafer fabrication, the term 'lift-off' refers to the process of creating patterns on the wafer surface through an additive process, as opposed to the more familiar patterning techniques that involve subtractive processes, such as etching. Lift-off is most commonly employed in patterning metal films for interconnections. Lift-off consists of forming an inverse image of the pattern desired on the wafer using a stencil layer, which covers certain areas on the wafer and exposes the rest. The layer to be 'patterned' is then deposited over the 'stenciled' wafer. In the exposed areas of the stencil, the layer material gets deposited directly on the wafer substrate, while in the covered areas, the material gets deposited on top of the stencil film. After the layer material has been deposited, the wafer is immersed in a liquid that can dissolve the stencil layer. Once the stencil is dissolved by the liquid, the layer material over it gets 'lifted off' (hence the term 'lift-off'), leaving behind the layer material that were deposited over the wafer substrate itself, which forms the final pattern on the wafer.

The 'lift-off' process as a patterning technique offers the following advantages: 1) composite layers consisting of several different materials may be deposited one material at a time and then 'patterned' with a single 'lift-off'; 2) residues that are difficult to remove are prevented in the absence of etching of the patterned layer; 3) sloped side walls become possible, resulting in good step coverage. On the other hand, the main disadvantage of the lift-off process is the difficulty of creating the required stencil patterns for successful lift-off. Materials that have already been used as stencil film for 'lift-off' include: 1) a single photoresist layer; 2) two photoresist layers; 3) a photoresist-aluminum-photoresist layer; 4) polyimide/molybdenum layer; 5) polyimide/polysulfone/SiO layer; and 6) inorganic dielectric-photoresist layer.

A technique to define the metals without using any etchant, known as metal lift-off. This technique is an alternative to the conventional wet etching process. In this technique, a metal pattern is defined on the photoresist before the metal is deposited on the wafer. Finally, solvent is used to remove the photoresist. The metal which sits on the photoresist

will be lifted off, together with the photoresist. This leave the remaining metal which is not located on top of the photoresist.

Figure 2.15 shows the lift-off process where (i) photolithography & metallization (ii) metal lift-off (iii) metal contact defined [4].The key to successful lift-off is the ability to ensure the existence of a distinct break between the layer material deposited on top of the stencil and the layer material deposited on top of the wafer substrate. Such a separation allows the dissolving liquid to reach and attack the stencil layer. One technique to create such 'breaks' is cold evaporation over steep steps.
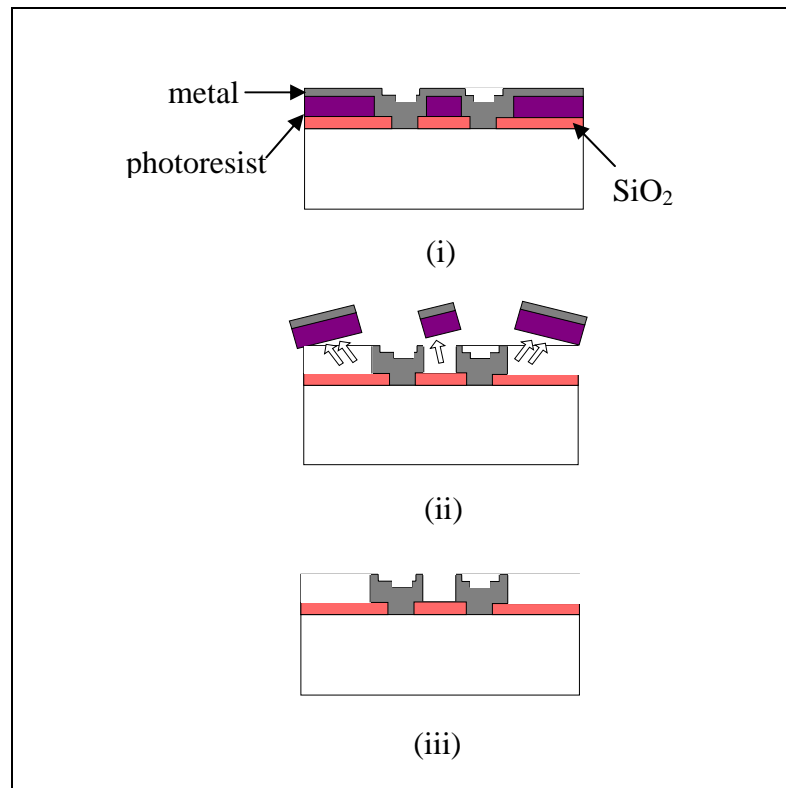

Figure 2.15. Wafer cross section for the lift-off process (i) photolithography & metallization (ii) lift-off (iii) metal contact defined

2.4   AlNiAu Process Using Electroless UBM Method

Electroless bumping process is the selective autocatalytic metal deposition on activated Al pads without any costly equipment such as plating base and lithography. The nickel, cobalt, palladium, platinum copper, silver and variety of alloys involving one of more of these metals can be deposited with this technology [5]. Electroless nickel bump is unique, because it acts as UBM that can be used as a foundation for solder bump or with slight thickness increase it acts as a stand alone minibump that can be used with conductive adhesives. This can be achieved because electroless nickel is isotropic [6], the bumps grew evenly in all directions, and when above the passivation level, spread over it, sealing the bond pads.

In electroless nickel bumping, the process starts with the wafer back side coating to prevent from nickel plated on the exposed silicon. The following next is cleaning process, used to remove contaminant on the passivation and Al bond pad. A second cleaning process is then applied to removes thick Al oxides and prepares the surface for metal deposition. After that is the zincation process. It activates the Al bond pads surface for Ni deposition. A thin zinc layer is deposited on Al which is substitutes by Ni in the Ni bath. Finally, a thin gold layer is deposited on the Ni from an immersion gold bath to prevent the oxidation of Ni before soldering and also help to improve solderability with solder ball.

2.4.1   Aluminum cleaning

Aluminum cleaning process is cleaned aluminum pads were micro-etched to remove soft aluminum oxide and activated for better nucleation at subsequent process an also to performed to remove various contaminations that were presents on the wafer surfaces that were due to material handling. The cleaning process can be done on many procedures with using many chemical including vapors process, emulsion, ultrasound, oxidation term, plasma recharge.
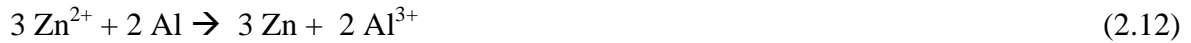
2.4.2 Aluminum activation

The purposes of activation process are to micro-etch the aluminum oxide and activate the surface for better nucleation at subsequent process and also necessary to remove native oxide from the aluminum surface and to form a layer which can initiate electroless nickel deposition. The controlling variables that also affect the oxide removal rate are the pH and temperature of the alkaline solution. The higher the pH, the faster the etch rate, and the higher the temperature the faster the etch rate. Tight control of these two parameters is essential in order to have a reproducible process and reliable product.

2.4.3 Zincation Activation

The zincation is a process of coating the aluminum bond pad with a layer of zinc. The zinc layer will prevent the re-oxidation on aluminum bond pad and to initiate the electroless nickel deposition. During the zincation process the aluminum oxide completely dissolved into solution while depositing zinc-nickel on the surface. The zincation process is usually repeated twice in order to produce a uniform layer of zinc on the surface of the aluminum. The important parameters related to zincation are: (1) time in the zincate solution, (2) time in the nitric acid zinc strip bath, (3) zinc concentration, (4) bath age, and (5) aluminum concentration.

The quality of the nickel layer has a direct relationship to the quality of the zinc layer, i.e. grain size, uniformity, and adhesion. A zincation pretreatment of the Al is essential to enhance the Al-Ni interfacial contact, acting as a sacrificial layer during the autocatalytic electroless nickel plating process [7]. To improve the adhesion and properties of the films, the pretreatment zincate process is used [8-12], which involves the exposure of the component to a solution containing zincate ($[Zn(OH)]$) ions. Generally, a highly alkaline solution is employed with the basic components being ZnO and NaOH.

improvement of the nucleation process has been clearly demonstrated [13]. Below is the chemistry equation of removing (2.11) and zincation deposition (2.12).

$$Al_2O_3 + 6NaOH \rightarrow 2Al(OH)_3 + 3Na_2O \qquad (2.11)$$

$$3\ Zn^{2+} + 2\ Al \rightarrow 3\ Zn + 2\ Al^{3+} \qquad (2.12)$$

Anodic dissolution of aluminum:

$$Al_{(s)} + 3\ OH^- \rightarrow Al\ (OH)_3 + 3\ e^- \qquad (2.13)$$

Cathodic deposition of zinc:

$$Zn\ (OH)^{2-} \rightarrow Zn^{2+} + 4\ OH^- \qquad (2.14)$$

$$Zn^{2+} + e^- \rightarrow Zn_{(s)} \qquad (2.15)$$

2.4.4 Electroless nickel.

Electroless deposition refers to the autocatalytic or chemical reduction of metal ions plated to a base substrate. It is also referred as chemical plating over an active surface by the chemical reduction in the absence of an external electric current [14-15]. In electroless nickel immersion gold (ENIG) process, the aluminum oxide is chemically etched and replaced with a layer of zinc [16, 17, and 18]. Nickel is then electrolessly plated over the zinc as a barrier and solderable layer while immersion gold is plated over the nickel as an oxidation barrier to improve it solderability. Two major considerations must be fulfilled in order to make implementation of an electroless Ni/Au process cost competitive: (a) the equipment must be capable of high volume production,and (b) the process must be reproducible. The critical control variables on the wafer processing side of the plating system include:

(a) process sequence (single and double zincation),

(b) bath processing times

(c) bath temperature

(d) bath level (automatic water replenishment to maintain chemical concentration)

Zinc that was used to initiate the reaction was dissolved in the electroless nickel solution during the reaction with ion zinc as following equation:

Reduction:

$$Ni^{2+} + 2e^- \rightarrow Ni \qquad (2.16)$$

Oxidation:

$$Zn \rightarrow Zn^{2+} + 2e^- \qquad (2.17)$$

General formulation of electroless Ni(P) deposition.

$$Ni^{2+} + H_2O + NaH_2PO_2 \rightarrow Ni(P) + H_2 + NaH_2PO_3B \qquad (2.18)$$

2.4.6   Immersion gold.

After completed electroless nickel process, the next following process is the deposition of gold layer was carried out to prevent the surface from oxidation and improves solderability. An immersion reaction is an oxidation reduction system in which an ion in solution is reduced to the metal at the expense of the surface metal, which is oxidized to an ion. The exchange occurs in one direction only. This is determined by the relative positions of the interacting metals in the electromotive force series (any metal higher in the electromotive series will displace from solution any metal below it in the series). In immersion plating of gold over nickel, gold ions from solution are reduced to gold metal. The electrons needed for this reduction are supplied by the nickel substrate itself. Immersion deposition or displacement plating will cease as soon as the substrate is completely covered by the immersion coating:

$$Ni + Au^{++} \rightarrow Ni^{++} + Au \qquad (2.19)$$

Reduction

$$Au^+ + e^- \rightarrow Au \qquad (2.20)$$

Oxidation

$$Ni \rightarrow Ni^{2+} + 2e^- \qquad (2.21)$$